

1
2 **IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

3 Application Serial No.09/843,255
4 Filing Date04/24/2001
5 First Named Inventor..... Venkatesan
Applicant Microsoft Corporation
6 Attorney's Docket No. MS1-647US
7 Title: *Recognizer of Text-Based Work*

8

Examiner	Phone	Fax	Group Art Unit
BALI VIKKRAM	(571)272-7415	(703) -	2623

9

10 **APPEAL BRIEF**

11 To: MS: Appeal Brief - Patents
12 Commissioner for Patents
13 P.O. Box 1450
Alexandria, VA 22313-1450

14 From: Kasey C. Christie (Tel. 509-324-9256; Fax 509-323-8979)
15 Customer No. 22801

16
17 Pursuant to 37 C.F.R. §1.192, Applicant hereby submits an appeal brief for
18 Application No. 09/843,255. A Notice of Appeal was filed October 14, 2005.
19 Accordingly, Applicant appeals to the Board of Patent Appeals and Interferences
20 seeking review of the Examiner's rejections.
21
22
23
24
25

TABLE OF CONTENTS

<u>Appeal Brief Items</u>	<u>Page</u>
(1) Real Party in Interest	3
(2) Related Appeals, Interferences, and Judicial Proceedings	3
(3) Status of Claims	4
(4) Status of Amendments	5
(5) Summary of Claimed Subject Matter	6
(6) Grounds of Rejection to be Reviewed on Appeal	8
(7) Argument	9
(8) Appendix of Appealed Claims	21

1 **(1) Real Party in Interest**

2 The real party in interest is the Microsoft Corporation, the assignee of all
3 right and title to the subject invention.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 **(2) Related Appeals, Interferences, and Judicial Proceedings**

2 Appellant is not aware of any other appeals, interferences, or judicial
3 proceedings which will directly affect, be directly affected by, or otherwise have a
4 bearing on the Board's decision to this pending appeal.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 **(3) Status of Claims**

2 Claims 1-7, 15-19, 64, 65, and 67-71 are pending in this Application, and
3 are set forth in the Appendix of Appealed Claims on page 21. All pending claims
4 (claims 1-7, 15-19, 64, 65, and 67-71) stand rejected. Claims 1-66 were originally
5 filed in the Application. Claims 8-14, 20-62, 63, and 66 have been canceled,
6 withdrawn, and/or non-elected. Claims 1, 7, 15, 64, and 65 have been amended.
7 In response to a restriction requirement, Applicant elected claims 1-7, 15-19, 64
8 and 65. Claims 67-71 were added. No claims have been allowed.

9 Claims 1-7, 15-19, 64, 65, and 67-71 are subject of this appeal and stand
10 rejected as set forth in a Final Office Action dated June 22, 2005 (hereinafter, the
11 "FINAL ACTION").

12 Specifically, as set forth in p. 2 of the FINAL ACTION, claims 1-7, 15-19,
13 64, 65, and 67-71 stand rejected under USC § 103(a) as being obvious in light of
14 the combination of Li: *Liang Li*, US Patent No. 5,774,588 (issued 06/30/1998) and
15 **Karaev: *Isaak Karaev et al***, US Patent No. 5,802,518 (issued 09/01/1998).

1 **(4) Status of Amendments**

2 The Applicant responded to a non-final Office Action issued on October
3 14, 2004 (hereinafter, the "NON-FINAL ACTION"). In that response, Applicant
4 only amended claims 1, 7, 15, 64, and 65 and added claims 67-71. Applicant
5 traversed all substantive rejections and amended no other claims.

6 After that, the FINAL ACTION issued on June 22, 2005—the action
7 dismissing Applicant's traversal and maintaining the rejection of all pending
8 claims. In Applicant's response to the FINAL ACTION, Applicant traversed all
9 substantive rejections and amended no claims. No other amendments have been
10 filed subsequent to the FINAL ACTION.

11 The Office issued an advisory action on September 7, 2005 (hereinafter, the
12 "ADVISORY ACTION"). No other amendments have been filed subsequent to
13 the FINAL ACTION or advisory action.

1 **(5) Summary of Claimed Subject Matter**

2 Broadly speaking, the claimed subject matter describes a technology for
3 recognizing the content of text documents. The technology may detect similarity
4 between text-based works in an automatic and accurate manner. Generally, the
5 technology determines one or more hash values for the content of a text document.
6 Furthermore, the technology may generate a “sifted text” version of a document.

7 In one implementation described in the application, document recognition
8 is used to determine whether the content of one document is copied (i.e.,
9 plagiarized) from another document. This is done by comparing hash values of
10 documents (or alternatively their sifted text).

11 Following is a concise explanation of each independent claim 1, 15, 64, and
12 65 involved in the Appeal which includes specification references and exemplary
13 drawing reference characters. As explained, the independent claims are not
14 limited solely to the elements identified by the reference characters.

15 Specifically:

16 Claim 1 includes obtaining (410) a body of text containing textual content
17 in a computer-readable format; formatting (418) the body of text into a defined
18 image-based format, wherein the textual content of the defined image-based
19 formatted body of text is immutable via software tools for manipulation of textual
20 content of bodies of text; deriving (420) a hash value representative of the textual
21 content of the body of text, perceptually distinct bodies of text having hash values
22 that are substantially independent of each other.

1 Claim 15 includes obtaining (410) a body of text containing textual content
2 in a computer-readable format; formatting (418) the body of text into a defined
3 image-based format, wherein the textual content of the defined image-based
4 formatted body of text is immutable via software tools for manipulation of textual
5 content of bodies of text; and deriving (420) a hash value representative of the
6 body of text, perceptually similar bodies of text having proximally similar hash
7 values.

8 Claim 64 includes obtaining (410) a body of text containing textual content
9 in a computer-readable format; formatting (418) the body of text into a defined
10 image-based format, wherein the textual content of the defined image-based
11 formatted body of text is immutable via software tools for manipulation of textual
12 content of bodies of text; and deriving (420) a hash value representative of the
13 textual content of the body of text, perceptually distinct bodies of text having hash
14 values that are substantially independent of each other.

15 Claim 65 includes obtaining (410) a body of text containing textual content
16 in a computer-readable format; formatting (418) the body of text into a defined
17 image-based format, wherein the textual content of the defined image-based
18 formatted body of text is immutable via software tools for manipulation of textual
19 content of bodies of text; and deriving (420) a hash value representative of the
20 body of text, perceptually similar bodies of text having proximally similar hash
21 values.
22
23
24
25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

(6) Grounds of Rejection to be Reviewed on Appeal

Whether claims 1-7, 15-19, 64, 65, and 67-71 are obvious under 35 U.S.C. § 103(a) in view of the combination of **Li** and **Karaev** and whether the Office has satisfactorily met its burden to show such obviousness?

1 (7) Argument

2
3 Issue -- Whether claims 1-7, 15-19, 64, 65, and 67-71 are obvious
4 under 35 U.S.C. § 103(a) in view of the combination of Li and Karaev and
5 whether the Office has satisfactorily met its burden to show such
6 obviousness?

7
8 Cited References

9 The Office cites Li as its primary reference in its obviousness-based
10 rejections. The Li reference is owned to the United Parcel Service, Inc. (UPS) and
11 is apparently utilized in electronically reading (e.g., via optical character
12 recognition) addresses on packages. Li describes a technology for efficiently
13 comparing an unverified string to a "lexicon," which filters the lexicon through
14 multiple steps to reduce the number of entries to be directly compared with the
15 unverified string.

16 The Li method begins by preparing the lexicon with an n-gram encoding,
17 partitioning and hashing process, which can be accomplished in advance of any
18 processing of unverified strings. The unknown is compared first by partitioning
19 and hashing it in the same way to reduce the lexicon in a computationally
20 inexpensive manner. This is followed by an encoded vector comparison step, and
21 finally by a direct string comparison step, which is the most computationally
22 expensive.

23 The reduction of the lexicon is accomplished without arbitrarily eliminating
24 any large portions of the lexicon that might contain relevant candidates. At the
25

1 same time, the method avoids the need to compare the unverified string directly or
2 indirectly with all the entries in the lexicon. The final candidate list includes only
3 highly possible and ranked candidates for the unverified string, and the size of the
4 final list is adjustable.

5 The Office also cites **Karaev** in its obviousness-based rejections. This
6 reference is owned by Multex.com, Inc. and describes the secure electronic
7 distribution of research documents over the World Wide Web to investors who are
8 authorized to receive said research documents. A repository server receives
9 research documents from contributors.

10 Also received are corresponding document profiles with information
11 relating to each research document including authorization information specifying
12 who is permitted to access each research document. The repository server includes
13 a first database for structured query searches and a second database for full text
14 searches.

15 A web server is coupled to the repository server and coupled to the World
16 Wide Web. The web server receives requests from investors for research
17 documents that satisfy a query. The web server determines whether the first
18 database or the second database should be searched based upon the type of query.
19 The repository server transmits to the web server a list of research documents that
20 satisfy the query and which the investor is authorized to access according to the
21 authorization information. The web server then formats the list of documents
22 according to a template form. The system also has a control mechanism to prevent
23 concurrent unauthorized access by two people using the same ID/password
24 combination.
25

Claim 1

With the cited portions of **Li** and **Karaev** provided in brackets, claim 1 recites (in part):

- obtaining a body of text containing textual content in a computer-readable format; [**Li**: Figs. 1A and 6, step 100; Fig. 1B, step 120; Col. 6, lines 40-50]
- formatting the body of text into a defined image-based format, wherein the textual content of the defined image-based formatted body of text is immutable via software tools for manipulation of textual content of bodies of text; [**Karaev et al.**: Col. 3, lines 28-30 and lines 33-37]
- deriving a hash value representative of the textual content of the body of text, perceptually distinct bodies of text having hash values that are substantially independent of each other. [**Li**: Figs. 2, 4A-B, and 5; col. 7, lines 17-67, col. 8, lines 1-14];

The Office admits that **Li** does not teach the “formatting” action, as recited in the claim. The Office therefore relies on **Karaev**, which teaches a remotely accessible web server that takes contributed documents and converts them to PDF format. On page 3 of the Action, the Office states that it would be obvious to combine the “teaching of [] **Karaev** of converting the data into the PDF format in order to safer guard the data, in to **Li**’s system.” Applicant, however, submits that

1 **Li** and **Karaev** do not teach or suggest all of the elements of claim 1, as there
2 exists no motivation to combine these references in this way.

3
4 *No Motivation to Combine References*

5 Applicant submits that there is no suggestion, teaching, or reason given by
6 one reference that would motivate one of ordinary skill in the art at the time of the
7 invention (hereinafter, "OOSA") to combine it with the teachings of the other
8 reference. More specifically, there is no motivation to combine because:

- 9 • the combination of references would result in an inoperable method;
- 10 • the combination of references destroys or frustrate the intended
11 function of **Li**;
- 12 • no motivation exists in the references themselves to make the
13 combination; and
- 14 • the Office has not provided any objective and particular evidence
15 showing why OOSA would be motivated to combine the teachings
16 of the two references.

17 Accordingly, the Applicant therefore respectfully asks the Office to
18 withdraw its rejection of these claims.

19
20 *No Motivation to Combine: Combination would result in an inoperable*
21 *method*

22 Applicant submits OOSA would have no motivation to combine the
23 teachings of **Karaev** with **Li** because the resulting device/approach would result in
24 an inoperable method. The **Li** method is utilized for electronically reading (e.g.,
25 via optical character recognition) addresses on packages. According to **Li**, this is

done by efficiently comparing an unverified string (e.g. an incorrect city name on a package) to a “lexicon” (e.g. a list of all U.S. cities). In order to do so, however, the unverified string of Li must be modified during the disclosed method. As Li states, “[a]n unverified string is processed by the partitioning and hashing subroutine of Figs. 2 and 4.” (Col. 8, lines 60-61). This must be done so that a “small subset of the lexicon which shares characteristics with the unverified string” can be found. (Col. 8, lines 53-54).

If, however, the proposed modification of formatting the string “into a defined image-based format” that is “immutable via software tools” is performed, the Li method could not modify (i.e. partition and/or hash) the string as is necessary to complete the disclosed process. Therefore, the proposed modification would result in an inoperable method. No motivation to combine exists where such a result occurs.

No Motivation to Combine: Combination would Destroy or Frustrate the Intended Purpose

Applicant submits OOSA would have no motivation to combine the teachings of Karaev with Li because the resulting device/approach will destroy or frustrate Li’s intended purpose. The stated purpose of Li is to “more efficiently” compare a text string to a lexicon of text strings. (Col. 14, lines 15-17) (emphasis added). The example provided by Li and cited by the Office illustrates this purpose of efficiency. (See Col. 13, line 66 through Col. 14, line 14). This example demonstrates that the Li comparison method can be completed in less than 54 milliseconds.

The Office indicates the combined teaching of the cited references suggests a modification of **Li** so that its text string is “a defined image-based format.” However, Applicant submits that this modification to **Li** results in a less efficient process rather than a more efficient process. Indeed, not only might this result in an inoperable device, but it also adds an additional step to the **Li** process; thus further slowing down the comparison.

Moreover, as illustrated by Fig. 6 (and described in its related text), one of the first actions taken in the **Li** method is the conversion (via OCR) of an address on a package *from an image to a text*. Again, it would be directly adverse to the purpose of **Li** to immediately convert this just-converted text back again into an image. Therefore, no motivation exists to make this modification, as it would result in a much slower method, if the method were still operable at all.

For the foregoing reasons, Applicant submits that the combination of cited references results in a modification which is directly adverse to the purpose of the **Li** method. Accordingly, OOSA would have no motivation to combine the teachings of cited references.

No Motivation to Combine: Cited References Express no Motivation to Combine

Applicant submits OOSA would have no motivation to combine the teachings of **Karaev** with **Li** because the neither reference expresses a reason to combine the teachings of these references, either explicitly or implicitly. At p. 3 of the Action, the Office suggests that the motivation to combine the teaching of these references is “in order to safer guard the data” utilized in **Li**’s approach.

However, Applicant submits that **Li** never teaches, discloses, suggests or hints at any need to safeguard its data. The problem solved by **Li** relates to identifying the intended destination city that may be incorrectly or illegibly written on a package to be mailed. Applicant submits that this problem does not imply a need to safeguard data. It also does not suggest utilizing an image-based format which avoids “manipulation of textual content,” as recited in claim 1.

In addition, **Karaev** does not disclose that its teaching could be utilized in a system for comparing strings of text (like what **Li** discloses).

For the foregoing reasons, Applicant submits that neither reference expresses a reason to combine the teachings of these references. Accordingly, OOSA would have no motivation to combine the teachings of cited references.

No Motivation to Combine: No Showing of Objective Evidence

Furthermore, Applicant respectfully submits that the Office has not met its burden in showing a motivation to combine **Li** and **Karaev**. More specifically, the Office has not identified any objective and particular evidence found in the cited references that show why OOSA would be motivated to combine the teachings of the two cited references.

The Office has not identified any specific portion of the cited references as being objective and particular evidence that would have motivated OOSA to look towards the teachings of the other to produce the combination of references that the Office proposes. Applicant respectfully submits that the Office cannot maintain this obviousness-based rejection without pointing out, with particularity, the specific portions of the cited references that would have motivated OOSA to

1 look towards the teachings of the other to produce the combination of references
2 that the Office proposes.

3 For the foregoing reasons, Applicant submits that the Office has not met its
4 burden in showing objective evidence to combine references. Accordingly, OOSA
5 would have no motivation to combine the teachings of cited references.

6 In sum, Applicant submits that there is no suggestion, teaching, or reason
7 given by one reference that would motivate one of ordinary skill in the art at the
8 time of the invention (hereinafter, "OOSA") to combine it with the teachings of
9 the other reference. More specifically, there is no motivation to combine because
10 the combination of references destroys the intended function of Li; no motivation
11 exists in the references themselves to make the combination; and the Office has
12 not provided any objective and particular evidence showing why OOSA would be
13 motivated to combine the teachings of the two references.

14 Accordingly, the Applicant therefore respectfully submits that the rejection
15 of this claim should be withdrawn.

16
17 Claims 2-7 and 67-69

18 These claims ultimately depend upon independent claim 1. As discussed
19 above, claim 1 is allowable.

20 In addition to its own merits, each of these dependent claims is allowable
21 for the same reasons that its base claim is allowable. Applicant submits that the
22 Office withdraw the rejection of each of these dependent claims because its base
23 claim is allowable.
24
25

Moreover, claim 67 recites, “indicating suspicion of plagiarism between the two bodies of text when the compared hash values of the two bodies of text substantially match.” Applicant submits that the Office has not identified this feature in either cited reference nor has it shown this feature would be part of a combination that would be obvious to OOSA.

Furthermore, claim 69 recites, “before formatting, the textual content of the body of text comprises multiple words and sentences and the derived hash value is representative of all of the textual content of the body of text.” Applicant submits that the Office has not identified this feature in either cited reference nor has it shown this feature would be part of a combination that would be obvious to OOSA.

Claim 15

In rejecting claim 15, the Office stated that it was rejected for arguments analogous to those presented in rejecting claim 1. If this is true, then Applicant submits that claim 15 is allowable for at least the same reasons as discussed above in regards to claim 1. Accordingly, Applicant asks the Office to withdraw its rejection of claim 15.

Claims 16-19, 70, and 71

These claims ultimately depend upon independent claim 15. As discussed above, claim 15 is allowable.

In addition to its own merits, each of these dependent claims is allowable for the same reasons that its base claim is allowable. Applicant submits that the

Office withdraw the rejection of each of these dependent claims because its base claim is allowable.

Moreover, claim 71 recites, “before formatting, the textual content of the body of text comprises multiple words and sentences and the derived hash value is representative of all of the textual content of the body of text.” Applicant submits that the Office has not identified this feature in either cited reference nor has it shown this feature would be part of a combination that would be obvious to OOSA.

Claim 64

In rejecting claim 64, the Office stated that it was rejected for arguments analogous to those presented in rejecting claim 1. If this is true, then Applicant submits that claim 64 is allowable for at least the same reasons as discussed above in regards to claim 1. Accordingly, Applicant asks the Office to withdraw its rejection of claims 64.

Claim 65

In rejecting claim 65, the Office stated that it was rejected for arguments analogous to those presented in rejecting claim 15. If this is true, then Applicant submits that claim 65 is allowable for at least the same reasons as discussed above in regards to claim 15. Accordingly, Applicant asks the Office to withdraw its rejection of claims 65.

1 **Conclusion**

2 Based upon the foregoing reasons, Applicant submits that claims 1-7, 15-
3 19, 64, 65, and 67-71 are not obvious under 35 U.S.C. § 103(a) in view of the
4 combination of **Li** and **Karaev** and that the Office has not satisfactorily met its
5 burden to show such obviousness.

6 Applicant respectfully requests that the rejections be overturned and that
7 the pending claims 1-7, 15-19, 64, 65, and 67-71 be allowed to issue.
8

9
10
11
12 Dated: DEC. 19, 2005

Respectfully Submitted,
DAVE MORASCH, 42, 905 for

13 By: 

Kasey C. Christie
Reg. No. 40559
(509) 324-9256 x232
kasey@leehayes.com
www.leehayes.com

1 **(8) Appendix of Appealed Claims**

2
3 1. **(PREVIOUSLY PRESENTED)** A computer-implemented
4 method for hashing a body of text, the method comprising:

5 obtaining a body of text containing textual content in a computer-readable
6 format;

7 formatting the body of text into a defined image-based format, wherein the
8 textual content of the defined image-based formatted body of text is immutable via
9 software tools for manipulation of textual content of bodies of text;

10 deriving a hash value representative of the textual content of the body of
11 text, perceptually distinct bodies of text having hash values that are substantially
12 independent of each other.

13
14 2. **(ORIGINAL)** A method as recited in claim 0, wherein
15 perceptually distinct bodies of text have hash values that are independent of each
16 other.

17
18 3. **(ORIGINAL)** A method as recited in claim 0 further
19 comprising comparing hash values of two bodies of text to determine if such
20 values match.

21
22 4. **(ORIGINAL)** A method as recited in claim 0 further
23 comprising comparing hash values of two bodies of text to determine if such
24 values substantially match.
25

1 **5. (ORIGINAL)** A method as recited in claim 4 further
2 comprising indicating whether such values substantially match.

3
4 **6. (ORIGINAL)** A computer comprising one or more computer-
5 readable media having computer-executable instructions that, when executed by
6 the computer, perform the method as recited in claim 0.

7
8 **7. (PREVIOUSLY PRESENTED)** A computer-readable
9 medium having computer-executable instructions that, when executed by a
10 computer, performs the method as recited in claim 3.

11
12
13 Claims 8-14 are **CANCELED**.

14
15
16 **15. (PREVIOUSLY PRESENTED)** A computer-implemented
17 method for hashing a body of text, the method comprising:

18 obtaining a body of text containing textual content in a computer-readable
19 format;

20 formatting the body of text into a defined image-based format, wherein the
21 textual content of the defined image-based formatted body of text is immutable via
22 software tools for manipulation of textual content of bodies of text;

23 deriving a hash value representative of the body of text, perceptually
24 similar bodies of text having proximally similar hash values.

1 **16. (ORIGINAL)** A method as recited in claim 15 further
2 comprising comparing hash value of a body of text to determine if such value is
3 proximally near hash values of a group of bodies of text having proximally
4 clustered hash values.

5
6 **17. (ORIGINAL)** A method as recited in claim 16 further
7 comprising grouping the body of text with the group of bodies of text if the hash
8 value of such body is proximally near the values of the group.

9
10 **18. (ORIGINAL)** A computer comprising one or more computer-
11 readable media having computer-executable instructions that, when executed by
12 the computer, perform the method as recited in claim 16.

13
14 **19. (ORIGINAL)** A computer-readable medium having computer-
15 executable instructions that, when executed by a computer, performs the method
16 as recited in claim 16.

17
18 Claims 20-63 are CANCELED.

19
20 **64. (PREVIOUSLY PRESENTED)** A computer-readable
21 medium having computer-executable instructions that, when executed by a
22 computer, performs the method comprising:
23 obtaining a body of text containing textual content in a computer-readable
24 format;

1 formatting the body of text into a defined image-based format, wherein the
2 textual content of the defined image-based formatted body of text is immutable via
3 software tools for manipulation of textual content of bodies of text;

4 deriving a hash value representative of the textual content of the body of
5 text, perceptually distinct bodies of text having hash values that are substantially
6 independent of each other.

7
8 **65. (PREVIOUSLY PRESENTED)** A computer-readable
9 medium having computer-executable instructions that, when executed by a
10 computer, performs the method comprising:

11 obtaining a body of text containing textual content in a computer-readable
12 format;

13 formatting the body of text into a defined image-based format, wherein the
14 textual content of the defined image-based formatted body of text is immutable via
15 software tools for manipulation of textual content of bodies of text;

16 deriving a hash value representative of the body of text, perceptually
17 similar bodies of text having proximally similar hash values.

18
19 **66. (CANCELED)**

20
21 **67. (PREVIOUSLY PRESENTED)** A method as recited in
22 claim 4 further comprising indicating suspicion of plagiarism between the two
23 bodies of text when the compared hash values of the two bodies of text
24 substantially match.

1 **68. (PREVIOUSLY PRESENTED)**

2 A method as recited in
3 claim 0, wherein, before formatting, the textual content of the body of text
4 comprises multiple words and sentences.

5 **69. (PREVIOUSLY PRESENTED)**

6 A method as recited in
7 claim 0, wherein, before formatting, the textual content of the body of text
8 comprises multiple words and sentences and the derived hash value is
9 representative of all of the textual content of the body of text.

10 **70. (PREVIOUSLY PRESENTED)**

11 A method as recited in
12 claim 15, wherein, before formatting, the textual content of the body of text
13 comprises multiple words and sentences.

14 **71. (PREVIOUSLY PRESENTED)**

15 A method as recited in
16 claim 15, wherein, before formatting, the textual content of the body of text
17 comprises multiple words and sentences and the derived hash value is
18 representative of all of the textual content of the body of text.